OxyGene User Guide v1.0

© B@SIC, UMR 6026 – June 2008

1. Introduction

1.1. Objectives

OxyGene is a Client-Server application, which aims at constructing the sub-systems of the genes involved in oxidative stress for all sequenced bacterial genomes.

The sequences of the oxidative stress genes identified from the literature were aligned and analysed so as to establish a set of signatures, stored in a database called OxyDB. A novel ontology of the genes of detoxification, based upon the set of signatures, is proposed in OxyGene and enriches the existing ontology by being more precise. The repertories for all sequenced bacterial genomes are then obtained using the *OxyGene Annotator* from strict pattern matching between the genomes and the genes signatures. A Graphical User Interface, presented in this guide, has been developed to ease the query and bioanalysis processes.

1.2. Graphical User Interface (GUI)

Not only does OxyGene supply the subsystems of oxidative stress genes, but it also seeks to facilitate the job of the biologist or bioanalyst by providing a certain number of potentialities, such as:

- Investigate the presence or absence of the oxidative stress (OxyDB) genes within the chosen sequenced genomes;
- Enumerate the sequenced genomes that possess an instance of a given oxidative-stress gene;
- View and save under *fasta* format all nucleic and proteic sequences of any such gene of any chosen genome;
- View and save tables showing all details of all homologs of the OxyDB genes found;
- Associate a *Function* confidence level with each OxyDB gene class, and an *Annotation* confidence level with each particular instance of such a gene, with regard to objective criteria;
- Identify already annotated genes, re-annotated genes (because of a new start), and genes found *de novo*; furthermore isolate fragments, frameshifts or pseudogenes — for precise definitions see (Thybert & al., 2008);
- View the detoxification subsystem of any sequenced bacterial genome; this subsystem is a subset of the reference detoxification subsystem, which displays all known enzymatic potentialities of all sequenced bacteria;
- Compare and view a combination of detoxification subsystems (to see what different bacteria have in common, what oxidative compounds some union of bacteria can detoxify, etc.);
- View and compare the localisation of the oxidative stress genes on the replicons of any sequenced bacterial genomes;

2. Installation & Launch of the Application

OxyGene is a client-server application, with the server installed and staying at *OUEST-genopole Bio-Informatics Platform,* keeping all needed pre-computed genomic data, while the OxyGene Client or GUI is a Java application which communicates with the server via web-services. The OxyGene Client needs to be downloaded on your computer and can be found on web site:

```
http://www.umr6026.univ-rennes1.fr/english/home/research/basic/software/
```

In order to run, OxyGene needs Java JRE 5 (or a more recent version). If not already installed on your machine, the latter can be downloaded at the following address:

```
http://java.sun.com/javase/downloads/index_jdk5.jsp
```

Once OxyGene has been downloaded, unzip the OxyGene.zip or OxyGene.tar.gz file by clicking on it, or by typing under Linux:

```
tar -xzvf OxyGene.tar.gz
```

An `OxyGene/` directory should appear. In order to launch OxyGene, no matter which platform, first go to the `OxyGene/` directory;

On Windows, simply double-click on file: `OxyGene.bat`
On Mac OS X, double-click on file: `OxyGene.command`
On Linux, double-click on file: `OxyGene.sh` or in a terminal window, type: `./OxyGene`

Depending on the requests submitted, the OxyGene Client may require large amounts of memory. By default, this is somewhat accounted for within the files above. However, if your computer has less than 1Go available RAM, please replace the `-Xmx1g` option inside the above corresponding file by `-Xmx128m`, `-Xmx256m` or `-Xmx512m` according to your system available RAM.

The application OxyGene launches; after a few seconds, you should see the following window appear:

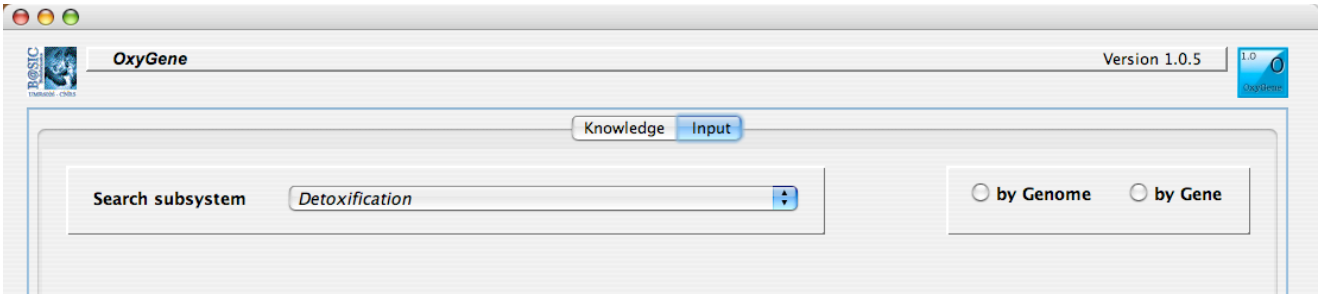


Figure 1. The Input tab at the beginning.

3. Description / Demonstration of the Application

At the beginning two different tabs appear in the OxyGene window:

- The Knowledge tab, which gives the *a priori* knowledge upon which OxyGene is based;
- The Input tab, which enables the biologist to express its request.

3.1. The Knowledge Tab

The *Knowledge* tab exhibits the *a priori* knowledge of OxyGene, represented in three sub-tabs:

(a) the taxonomy of genomes imported from Genbank,
(b) our OxyDB ontology of the oxidative stress genes (Figure 2),
(c) the reference subsystems which show all possible paths involved in oxidative stress used by all bacteria, with the OxyDB genes involved in each path (Figure 3);

The OxyDB genes are denoted by two identifiers: an *OxyDB number*, which draws inspiration — while being different — from the nomenclature of Margaret Riley, and an *OxyDB name*, e.g. CAT_GAT that stands for mono-functional catalase with Gatase domain, which we found were necessary to denominate the OxyDB genes in a more eloquent or expressive manner. For further details, please see the OxyDB documentation (OxyDB, 2008).

These three sub-tabs present some information for reference purposes, allowing the bio-analyst to answer such questions as: is our bacteria pre-computed in this OxyGene version? What is the OxyGene ontology? What are the reference subsystems — i.e. subsystems representing all enzymatic potentialities of all sequenced bacteria — of each domain of oxidative stress? These three sub-tabs are in no way intended to respond to user commands.

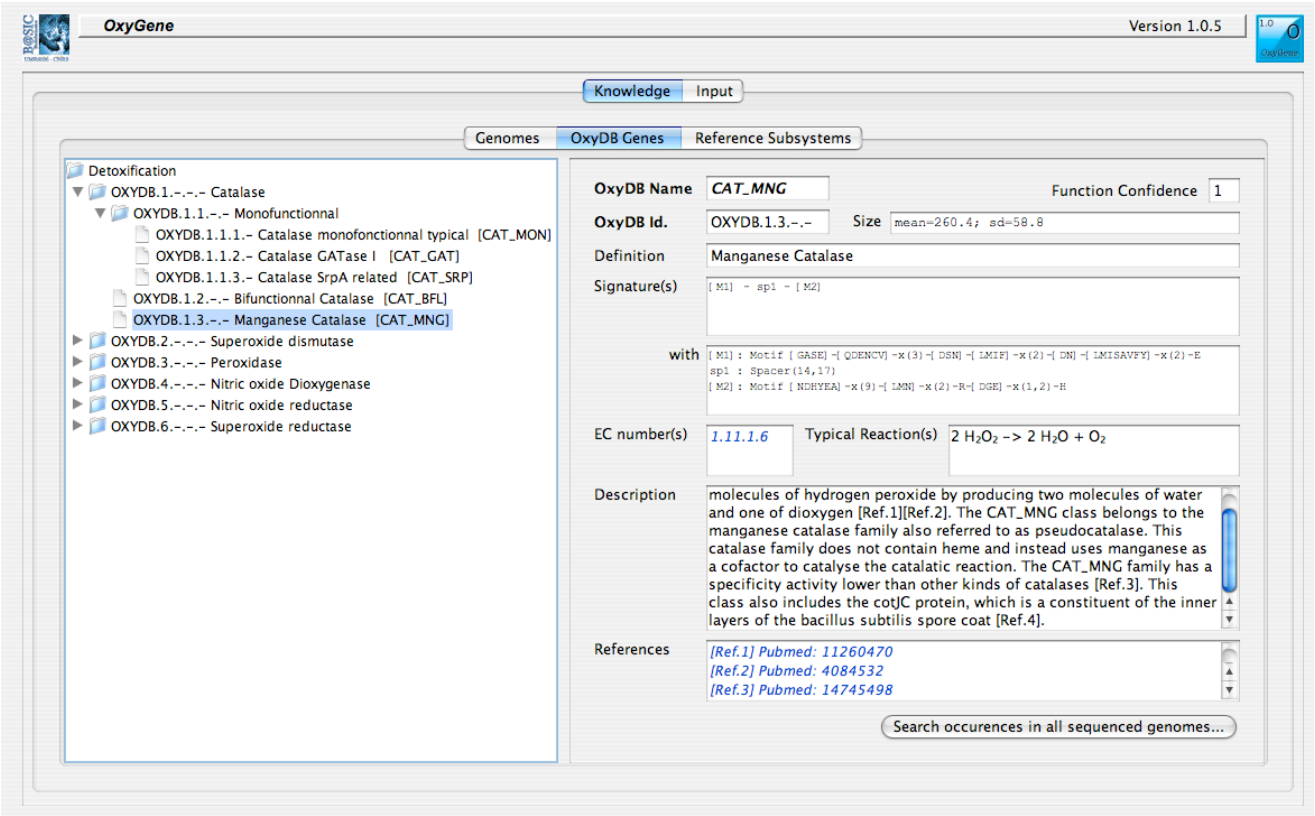

OxyGene
Version 1.0.5
Knowledge
Input
Genomes
OxyDB Genes
Reference Subsystems
Detoxification
OXYDB.1.-.-.- Catalase
OXYDB.1.1.-.- Monofunctionnal
OXYDB.1.1.1.- Catalase monofonctionnal typical [CAT_MON]
OXYDB.1.1.2.- Catalase GATase I [CAT_GAT]
OXYDB.1.1.3.- Catalase SrpA related [CAT_SRP]
OXYDB.1.2.-.- Bifunctionnal Catalase [CAT_BFL]
OXYDB.1.3.-.- Manganese Catalase [CAT_MNG]
OXYDB.2.-.-.- Superoxide dismutase
OXYDB.3.-.-.- Peroxidase
OXYDB.4.-.-.- Nitric oxide Dioxygenase
OXYDB.5.-.-.- Nitric oxide reductase
OXYDB.6.-.-.- Superoxide reductase
OxyDB Name
CAT_MNG
Function Confidence
1
OxyDB Id.
OXYDB.1.3.-.-
Size
mean=260.4; sd=58.8
Definition
Manganese Catalase
Signature(s)
[M1] - sp1 - [M2]
with
[M1] : Motif [GASE] -[QDENCV] -x(3) -[DSN] -[LMIF] -x(2) -[DN] -[LMISAVFY] -x(2) -E
sp1 : Spacer(14,17)
[M2] : Motif [NDHYEA] -x(9) -[LMN] -x(2) -R-[DGE] -x(1,2) -H
EC number(s)
1.11.1.6
Typical Reaction(s)
2 H2O2 -> 2 H2O + O2
Description
molecules of hydrogen peroxide by producing two molecules of water and one of dioxygen [Ref.1][Ref.2]. The CAT_MNG class belongs to the manganese catalase family also referred to as pseudocatalase. This catalase family does not contain heme and instead uses manganese as a cofactor to catalyse the catalatic reaction. The CAT_MNG family has a specificity activity lower than other kinds of catalases [Ref.3]. This class also includes the cotJC protein, which is a constituent of the inner layers of the bacillus subtilis spore coat [Ref.4].
References
[Ref.1] Pubmed: 11260470
[Ref.2] Pubmed: 4084532
[Ref.3] Pubmed: 14745498
Search occurences in all sequenced genomes...

Figure 2. OxyDB genes ontology and characteristics of each OxyDB gene.

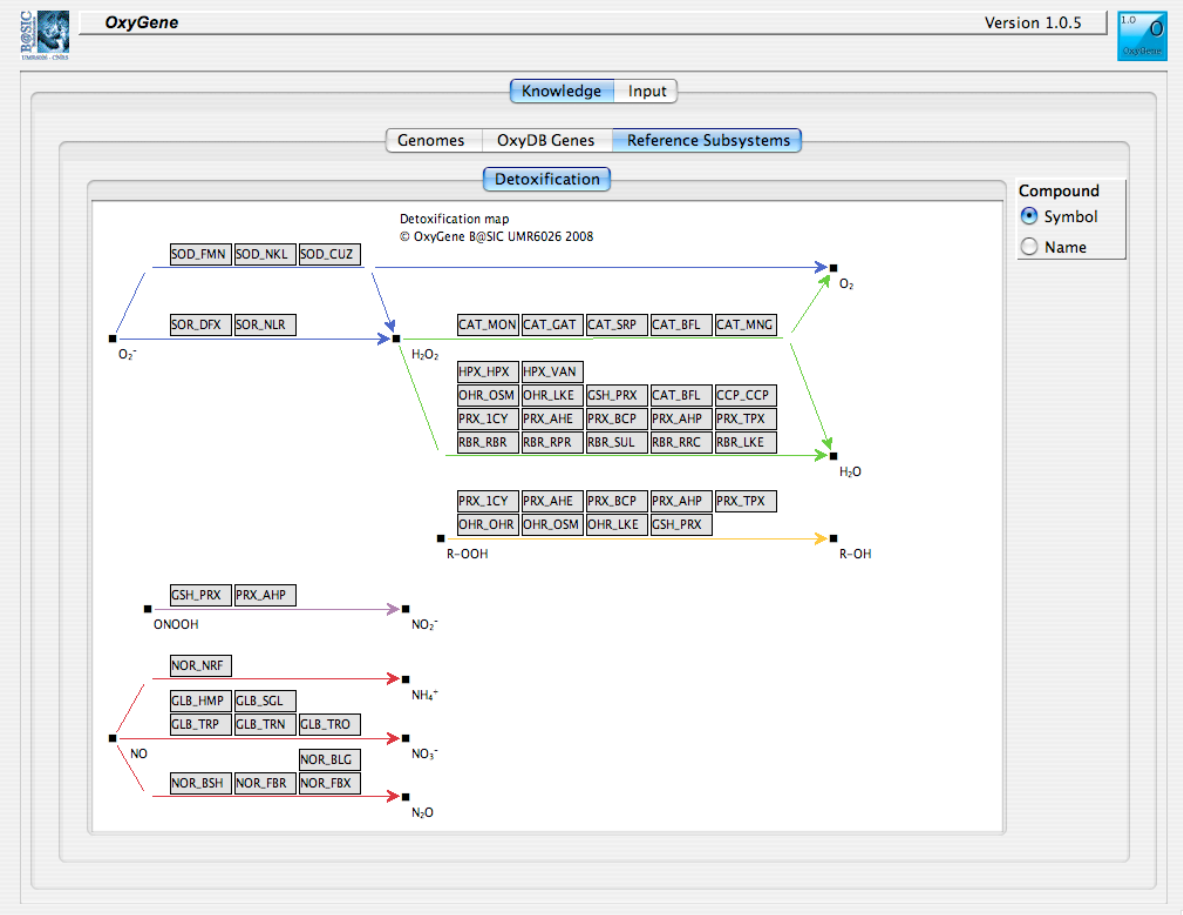


Figure 3. Detoxification reference sub-system, showing all known detoxification potentialities of all sequenced bacteria.

3.2. The Input Tab

The Input tab, already shown in Figure 1, allows the biologist to express its request. The *subsystem* — detoxification, reparation, reduction, regulation, etc. — must first be selected (only detoxification is available for the time being):

Then the biologist may submit two different kinds of requests to OxyGene:

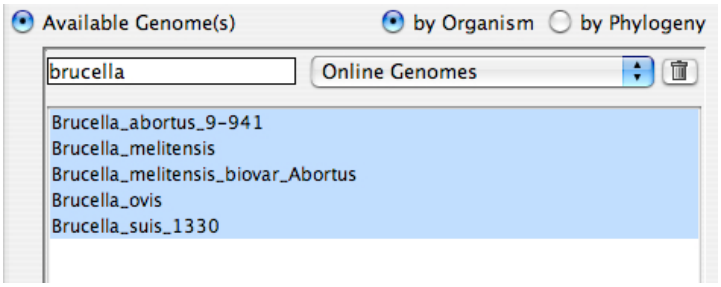
a) What are the oxidative stress genes that are present in some given genomes?
b) What are the genomes that contain an instance of a given OxyDB gene?

Depending on the question, the bio-analyst needs to check the appropriate radio button:

Question a): Select *Search by Genome*;
Question b): Select *Search by Gene*;

3.2.1. Search by Genome

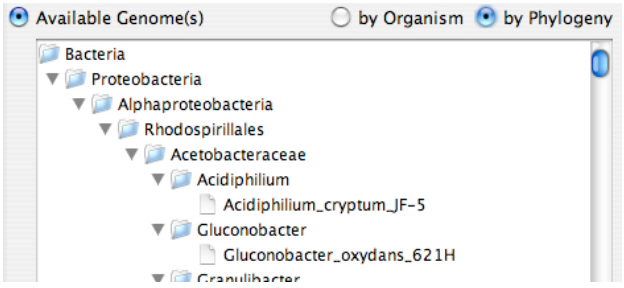
The genomes within which the biologist wants to search for OxyDB genes must then be selected. In this task, the biologist can be aided by using the editable field brucella to enter parts of the genome names:



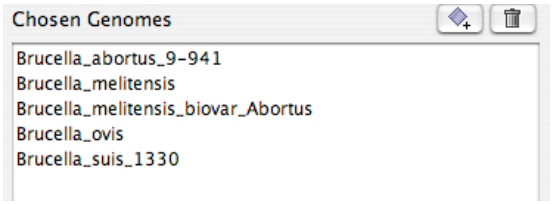
or by using the combo-box Online Genomes where groups of pre-selected genomes can be defined — see later. The button can be used to remove one such group.

The biologist may use the Shift+Click (interval selection) or Ctrl+Click combinations (disjoint selection) — +Click for Mac users — to select several genomes at once.

The genomes may also be selected from their phylogenetic group by clicking on the corresponding *by Phylogeny* radio button:



Click on the *Add* button Add > to add the selected genomes to the *Chosen Genomes* list box:



Genomes may be removed from this list box by selecting them and clicking on the button. A group of genomes containing all present genome names may also be defined here by clicking on the *Create new group of genomes* button ; a new editable field will appear Enter Group Name — simply enter a name for this new group of genomes and press *Enter*: the group will be added to the *Group of genomes* combo-box Online Genomes on the left-hand side, and may later be used to retrieve the selected genomes easily.

Once all the genomes to be analysed with respect to the oxidative stress repertory are imported in the list box *Chosen genomes*, the biologist may submit the request to OxyGene. This is performed by clicking on the **Determine Genes involved in Chosen Genomes** button. The OxyGene server then receives the request, reads it, recognizes the submitted names of the pre-computed genomes, and returns to the client the desired

data. The latter contains all the information regarding the selected genomes response to oxidative stress: i.e. which OxyDB genes are effectively present in these genomes, with their properties (locus tag, begin, end positions, confidence levels, sequence, etc.).

Once the data has returned from the server, the OxyGene client Window switches to present a table showing the presence or absence — the number of paralogs rather — of all OxyDB genes in the selected genomes (Figure 4):

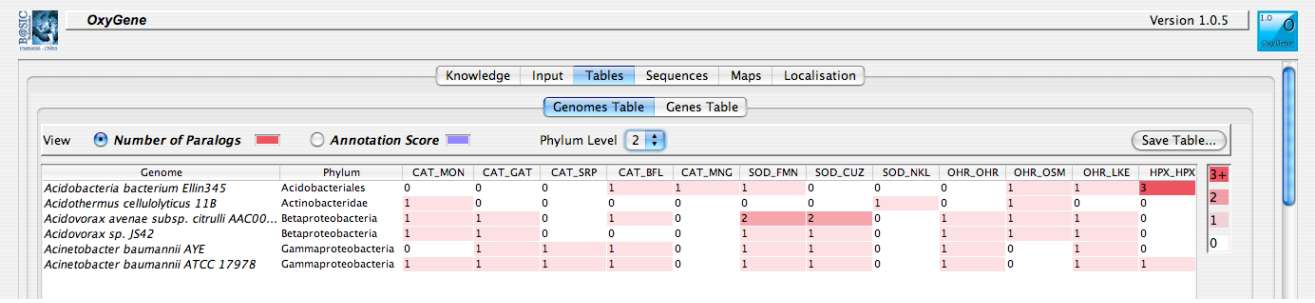


Figure 4. Submission of the chosen genomes to OxyGene results in the *Number of Paralogs* Table.

It can be seen that, in addition to the Knowledge and Input tabs, new tabs have been included in the OxyGene client window:

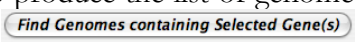
- A *Tables* tab showing different tables with lots of information (see Section 4.3.);
- A *Sequences* tab enabling the retrieval of all actual sequences of all OxyDB gene instances of all selected genomes;
- A *Maps* tab showing the subsystem (e.g. detoxification) for each selected bacterial genome; this subsystem will be a subset of the reference subsystem.
- A *Localisation* tab enabling the representation of the location of each OxyDB gene on all replicons of all selected genomes;

All these tabs will be described in details in Sections 3.3. to 3.6.

3.2.2. Search by Gene

The bio-analyst can investigate all genomes that possess a particular OxyDB gene or a combination of these by clicking on the *Search by Gene* radio button

. A new input interface then appears in the *Input* tab.

The OxyDB ontology appears on the left-hand side of the window. The biologist may now select one or more OxyDB genes and ask OxyGene to produce the list of genomes that contains at least one of those genes. This is achieved by clicking on the 

button.

As a result, the window presents the corresponding genomes (i.e. those with at least one selected OxyDB gene present) in alphabetical order in the upper list box, and the remaining genome names in the lower one (see Figure 5). By default, the corresponding genomes are selected and the remaining ones are not. It is possible to save the two lists in a text file (.txt) by clicking on the save button above the lists.

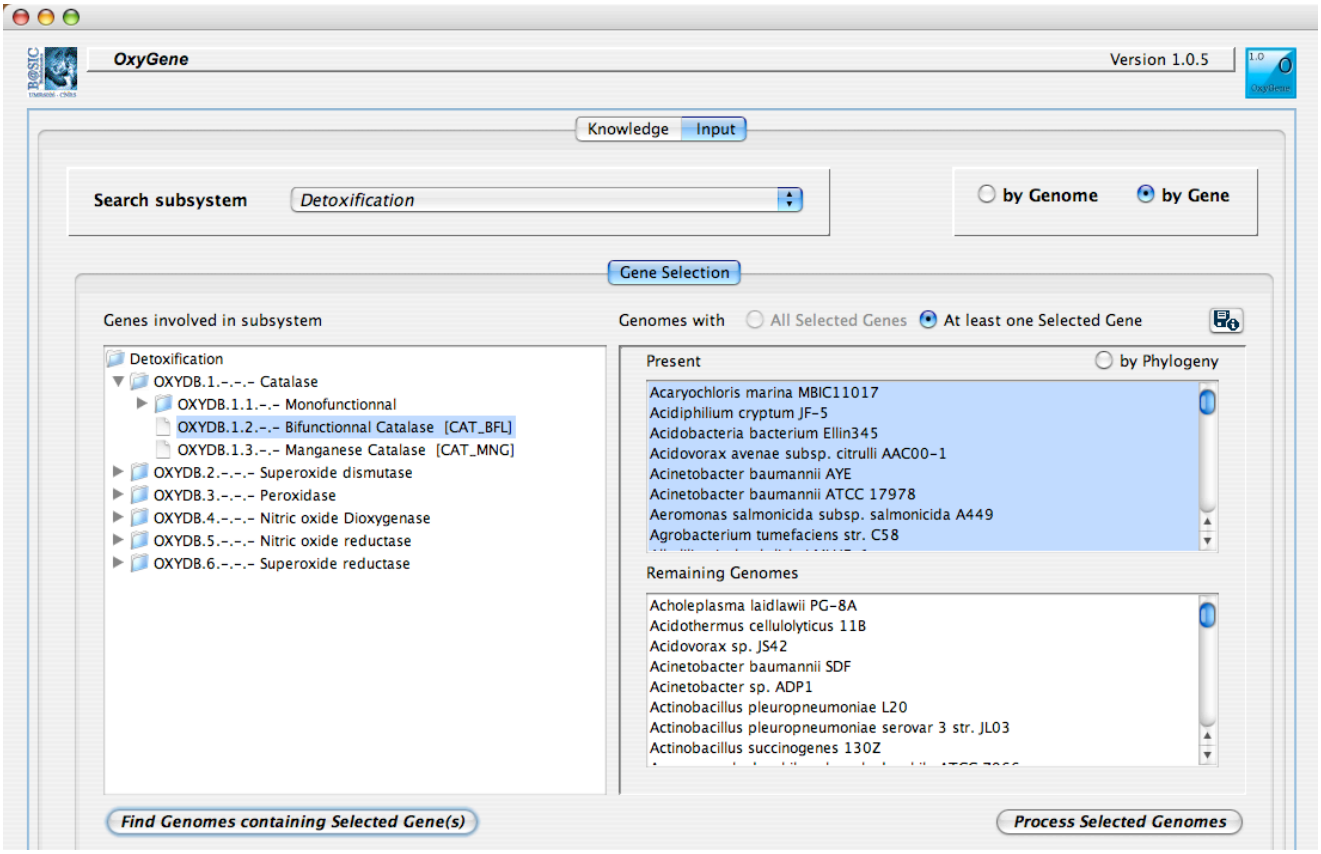
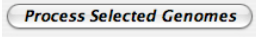
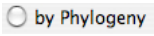


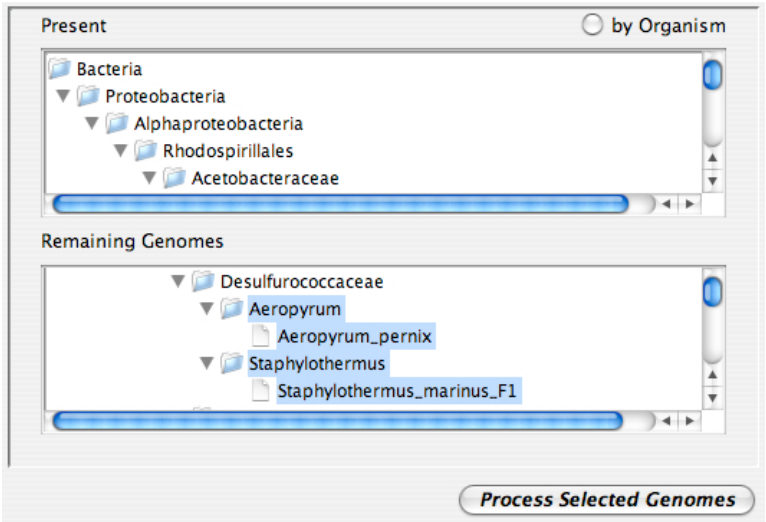
Figure 5. The resulting window when an OxyDB gene is selected and the '*Find Genomes…*' button pressed.

The biologist may then select/unselect genomes in the standard manner from the two list boxes OxyGene may then be asked to process the selected genomes by clicking on the button 

. This will send the request to OxyGene as usual, and will result in the appearance of the *Tables*, *Sequences*, *Maps* and *Localisation* tabs in the client window (as in Figure 4), representing all the information gathered on those selected genomes.

Finally, the radio button *by Phylogeny* 

at the top right corner of the two lists enables the biologist to select/unselect the genome names with respect to phylogenetic families:



3.3. The Tables Tab

Once some genome names have been selected, sent to the OxyGene server and processed, the client receives the available information from the server and displays it in tables shown in the *Tables* tab of the client window. There are two kinds of tables: the *Genomes* table and the *Genes* table.

3.3.1. The Genomes Table

The first table shown is the *Genomes Table*, which by default displays the *Number of Paralogs* table presenting the number of paralogs for each OxyDB gene for every submitted genome (see Figure 6). The different shades of the same colour represent different numbers of paralogs. The legend on the right hand side of the table indicates the colours used to represent each number.

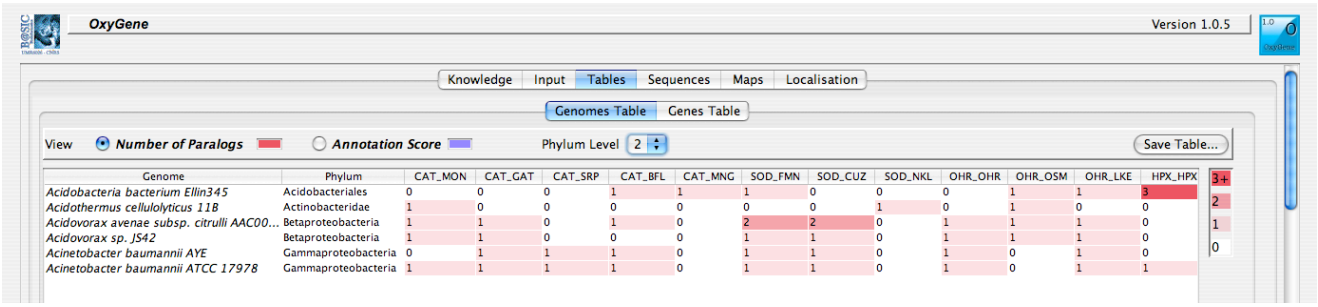


Figure 6. The *Number of Paralogs* table shows the number of paralogs for all OxyDB genes for all submitted genomes with different shades of the same colour to represent different numbers.

The radio-buttons allows the biologist to switch to the *Annotation Confidence Levels* table, and back. The latter confidence levels, shown in Figure 7, presents all different paralogs, identified by their locus tag, for all OxyDB genes for all submitted genomes in a single table, with different colours to indicate the annotation confidence levels associated to each particular instance of an OxyDB gene.

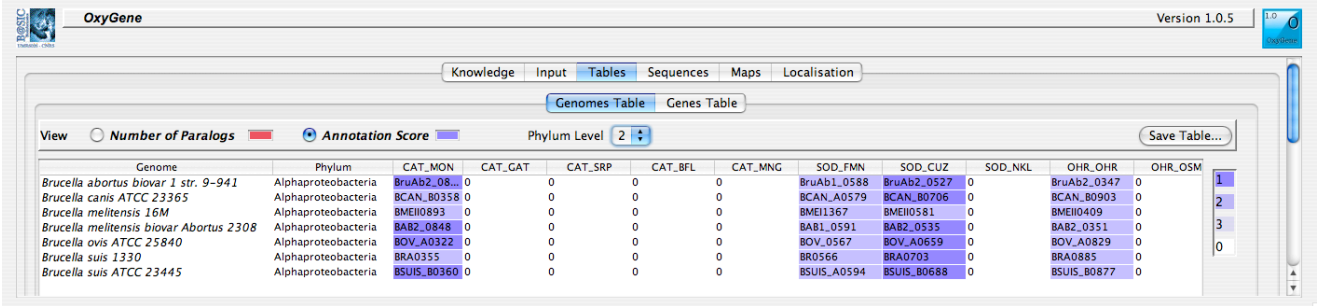


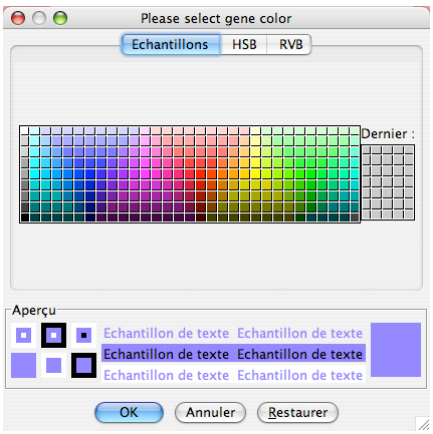
Figure 7. The *Annotation Confidence Levels* table shows all instances of all OxyDB genes for all submitted genomes, identifying them by their locus tag, and representing their annotation confidence levels by different shades of the same colour — here, all instances of the genes have the same annotation confidence level, i.e. 2.

The semantics of the *Annotation Confidence Level* — see (Thybert et al., 2008) — is the following:

3. The gene expression of this particular instance of the OxyDB gene has been experimentally verified, i.e. protein or RNA effectively expressed.
2. This particular instance of the gene belongs to an OxyDB group that contains at least one experimentally verified expressed gene.
1. This particular instance of the OxyDB gene seems to be ill-formed, i.e. a pseudogene, frameshift, or fragment.

The second column of the different tables is devoted to providing Phylum information; the biologist needs only specify the phylum level he/she wants to include in the table. This is achieved by selecting the appropriate level on the spin-box widget Phylum Level 2.

It is possible to change the fundamental colour, from which all the shades are derived, either for the *number of paralogs* colour or the *annotation confidence level* colour, by clicking on the colour button next to the corresponding radio-button. A colour definition window will then open:



Both tables can be saved as .txt or .xls files by clicking on the Save Table... button and specifying a repertory and name for the file. These files may ultimately be opened using Microsoft Excel or your favourite text editor.

Clicking on a coloured item (intersection of an OxyDB gene column and genome row) produces the opening of a new window, which provides detailed information to the biologist (see Figure 8).

The upper part of this window is the description of the OxyDB gene *class*, similar to that displayed in the OxyDB ontology of the *Knowledge* tab. The information provided in this part of the window is common to all instances. This part defines the OxyDB gene class and remains identical in the whole OxyDB column, specifying the OxyDB number, OxyDB name, corresponding EC numbers, function confidence level, its signature motifs, and providing a short description of the gene together with some references and links to related web sites.

The *Function Confidence Level* is attached to the OxyDB class and represents the degree to which its function may be trusted. Its semantics — again, see (Thybert et al., 2008) — is the following:

3. The enzymatic activity has been defined *in vitro* or *in vivo*.
2. The mutant phenotype suggested the gene function.
1. No function is defined, but its phylogenetic group is close to an OxyDB class, whose confidence level is 2 or 3.

The lower part enumerates the different *instances* of the OxyDB gene class in the considered genome. A different tab is created here for every instance of the gene; each tab specifies the genome and replicon in which the instance is found, its locus tag, its annotation usual name, its begin and end positions, its annotation confidence level and finally its specific sequence. As an exercise, one could check that the signature motifs can be found within the sequence.

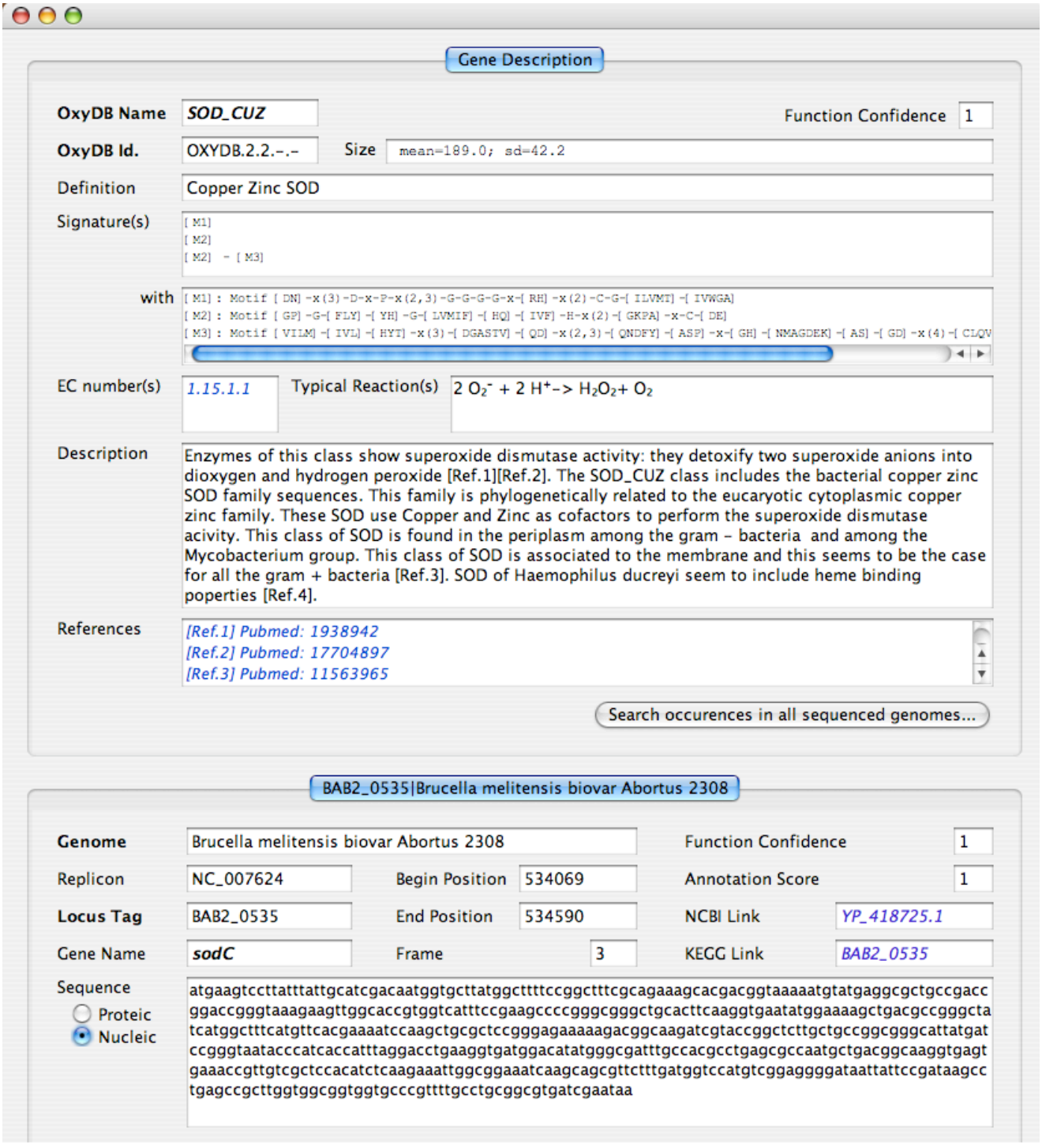

Gene Description
OxyDB Name
SOD_CUZ
Function Confidence
1
OxyDB Id.
OXYDB.2.2.-.-
Size
mean=189.0; sd=42.2
Definition
Copper Zinc SOD
Signature(s)
[M1]
[M2]
[M2] - [M3]
with
[M1] : Motif [DN] -x(3)-D-x-P-x(2,3)-G-G-G-G-x-[RH] -x(2)-C-G-[ILVMT] -[IVWGA]
[M2] : Motif [GP] -G-[FLY] -[YH] -G-[LVMIF] -[HQ] -[IVF] -H-x(2)-[GKPA] -x-C-[DE]
[M3] : Motif [VILM] -[IVL] -[HYT] -x(3)-[DGASTV] -[QD] -x(2,3)-[QNDFY] -[ASP] -x-[GH] -[NMAGDEK] -[AS] -[GD] -x(4)-[CLQV
EC number(s)
1.15.1.1
Typical Reaction(s)
2 O2- + 2 H+-> H2O2+ O2
Description
Enzymes of this class show superoxide dismutase activity: they detoxify two superoxide anions into dioxygen and hydrogen peroxide [Ref.1][Ref.2]. The SOD_CUZ class includes the bacterial copper zinc SOD family sequences. This family is phylogenetically related to the eucaryotic cytoplasmic copper zinc family. These SOD use Copper and Zinc as cofactors to perform the superoxide dismutase acivity. This class of SOD is found in the periplasm among the gram - bacteria and among the Mycobacterium group. This class of SOD is associated to the membrane and this seems to be the case for all the gram + bacteria [Ref.3]. SOD of Haemophilus ducreyi seem to include heme binding poperties [Ref.4].
References
[Ref.1] Pubmed: 1938942
[Ref.2] Pubmed: 17704897
[Ref.3] Pubmed: 11563965
Search occurences in all sequenced genomes...
BAB2_0535|Brucella melitensis biovar Abortus 2308
Genome
Brucella melitensis biovar Abortus 2308
Function Confidence
1
Replicon
NC_007624
Begin Position
534069
Annotation Score
1
Locus Tag
BAB2_0535
End Position
534590
NCBI Link
YP_418725.1
Gene Name
sodC
Frame
3
KEGG Link
BAB2_0535
Sequence
Proteic
Nucleic
atgaagtccttatttattgcatcgacaatggtgcttatggcttttccggctttcgcagaaagcacgacggtaaaaatgtatgaggcgctgccgacc
ggaccgggtaaagaagttggcaccgtggtcatttccgaagccccgggcgggctgcacttcaaggtgaatatggaaaagctgacgccgggcta
tcatggctttcatgttcacgaaaatccaagctgcgctccgggagaaaaagacggcaagatcgtaccggctcttgctgccggcgggcattatgat
ccgggtaatacccatcaccatttaggacctgaaggtgatggacatatgggcgatttgccacgcctgagcgccaatgctgacggcaaggtgagt
gaaaccgttgtcgctccacatctcaagaaattggcggaaatcaagcagcgttctttgatggtccatgtcggaggggataattattccgataagcc
tgagccgcttggtggcggtggtgcccgttttgcctgcggcgtgatcgaataa

Figure 8. OxyDB gene description window, showing the OxyDB gene class in its upper part, and the particular instances under different tabs in the lower part.

From this window, it is also possible to search for all occurrences of this OxyDB gene class in all sequenced bacterial genomes. Just click on the Search all occurences... button on the upper part of the window to achieve this: a new client window will open, automatically request all genomes containing that OxyDB gene from the server (*Search by Gene* request), process all these genomes, and display the *Number of Paralogs* table on those genomes (see Figure 5). It can be checked then that the column corresponding to that gene (i.e. for all submitted genomes) always contains at least one instance of that gene.

3.3.2. The Genes Table

Another table which is available within OxyGene is the *Genes* table, which recapitulates in a single table all information concerning the specific instances of the OxyDB genes found. The table presents for each genome, all its replicons, and for each replicon, all the OxyDB gene instances found, together with their OxyDB name, number, annotation confidence level, usual annotation name, locus tag, begin & end positions, frame, NCBI and KEGG web references. The Genes Tables is shown in Figure 9.

The latter two properties are written in blue italic and are clickable: this action will open your favourite web browser on the corresponding page (*NCBI* or *KEGG* page of the considered gene).

Finally, a *Save Table* button

enables the user to save the table under text (.txt) or Microsoft Excel (.xls) files.

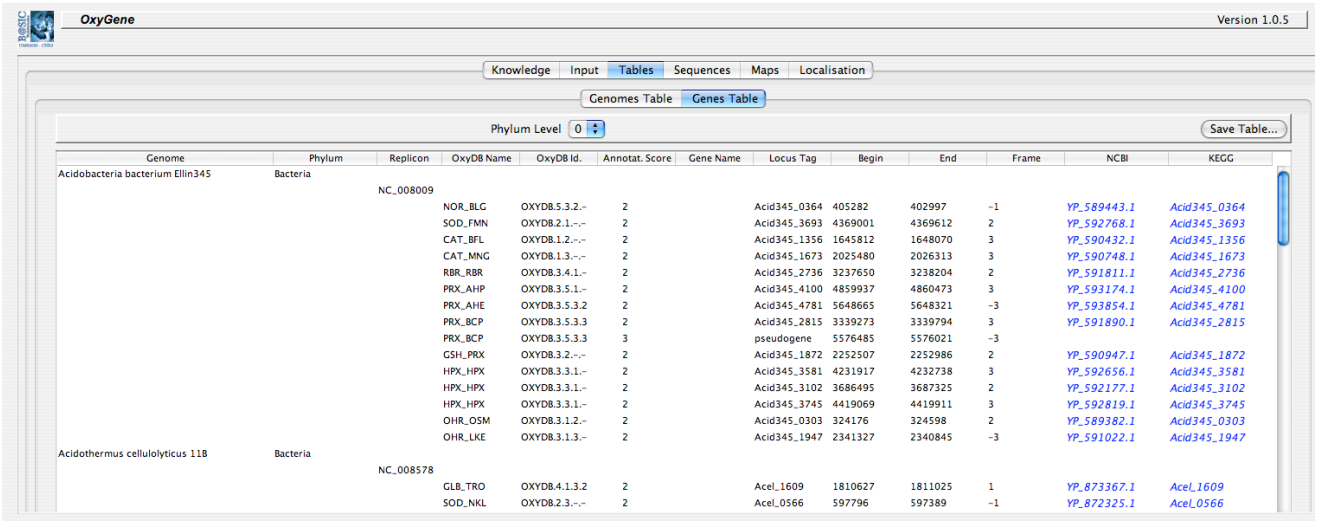
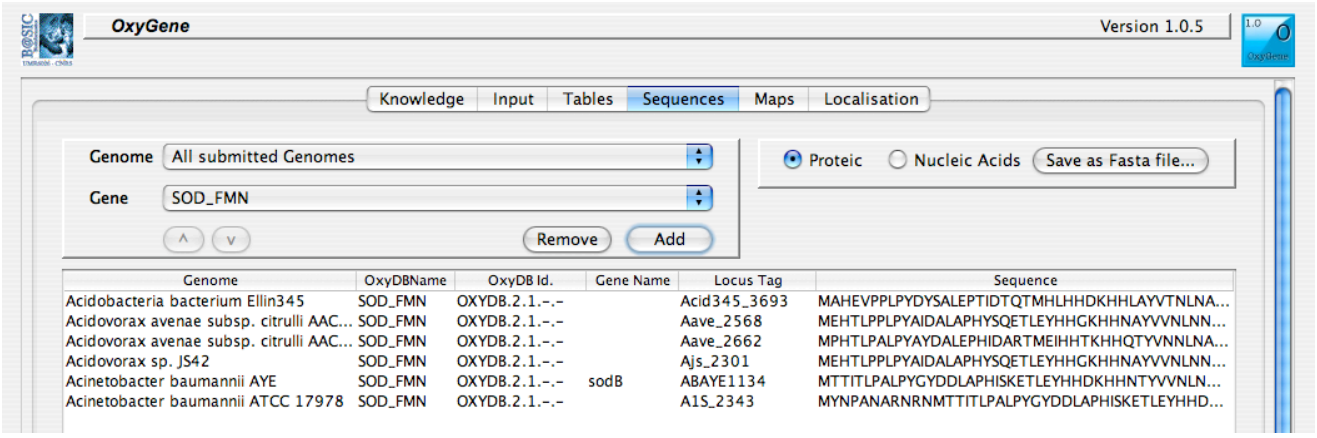
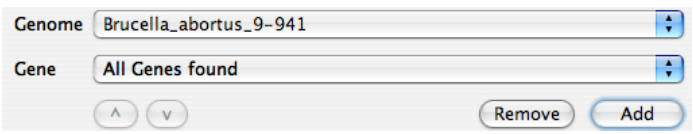


Figure 9. The Genes Table.

3.4. The Sequences tab


OxyGene
Version 1.0.5
Knowledge Input Tables Sequences Maps Localisation
Genome All submitted Genomes
Gene SOD_FMN
Remove Add
Proteic Nucleic Acids Save as Fasta file...
Genome OxyDBName OxyDB Id. Gene Name Locus Tag Sequence
Acidobacteria bacterium Ellin345 SOD_FMN OXYDB.2.1.-.- Acid345_3693 MAHEVPPLPYDYSALEPTIDTQTMHLHHDKHHLAYVTNLNA...
Acidovorax avenae subsp. citrulli AAC... SOD_FMN OXYDB.2.1.-.- Aave_2568 MEHTLPPLPYAIDALAPHYSQETLEYHHGKHHNAYVVNLNN...
Acidovorax avenae subsp. citrulli AAC... SOD_FMN OXYDB.2.1.-.- Aave_2662 MPHTLPALPYAYDALEPHIDARTMEIHHTKHHQTYVNNLNA...
Acidovorax sp. JS42 SOD_FMN OXYDB.2.1.-.- Ajs_2301 MEHTLPPLPYAIDALAPHYSQETLEYHHGKHHNAYVVNLNN...
Acinetobacter baumannii AYE SOD_FMN OXYDB.2.1.-.- sodB ABAYE1134 MTTITLPALPYGYDDLAPHISKETLEYHHDKHHNTYVVNLN...
Acinetobacter baumannii ATCC 17978 SOD_FMN OXYDB.2.1.-.- A1S_2343 MYNPANARNRNMTTITLPALPYGYDDLAPHISKETLEYHHD...

allows the biologist to easily reassemble in a fasta file the sequences of the desired genes. To do so, using the *Genome/Gene* selection panel:



the biologist must choose the desired genome — or may choose all genomes altogether —, then one of the OxyDB genes that exist in the selected genome — or all its OxyDB genes altogether —, then press the *Add* button to view the sequences. For each instance of the selected OxyDB gene class in the selected genome, a line will be added to the table, showing the genome name, OxyDB gene name & number, its usual annotation gene name, locus tag and the actual sequence of the gene instance.

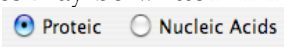
If *All submitted Genomes* and *All Genes found* are respectively selected in the *Genome* and *Gene* combo-boxes, then all instances of all OxyDB genes of all submitted genomes will be added to the table.

If *All submitted Genomes* is selected, and a single particular OxyDB gene class is selected, then all instances of that gene in all submitted genomes, i.e. all paralogs and orthologs, will be added to the table. This option enables the biologist to compare the different sequences implementing a particular OxyDB gene.

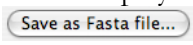
If a single particular genome is selected, and *All Genes found* is selected, then all instances of all OxyDB genes of that genome will be added to the table.

If a single particular genome is selected, and a single particular OxyDB gene class is selected, then all instances of that gene in that genome, i.e. all paralogs, will be added to the table.

To remove lines from the table, the user simply needs to select the desired lines within the table in the usual manner, and then press the *remove* button.

The sequences may be written in nucleic or proteic form. To switch form, simply select the appropriate radio-button 

.

All sequences displayed in the table may be saved into a fasta file, by clicking on the *Save as Fasta File* button 

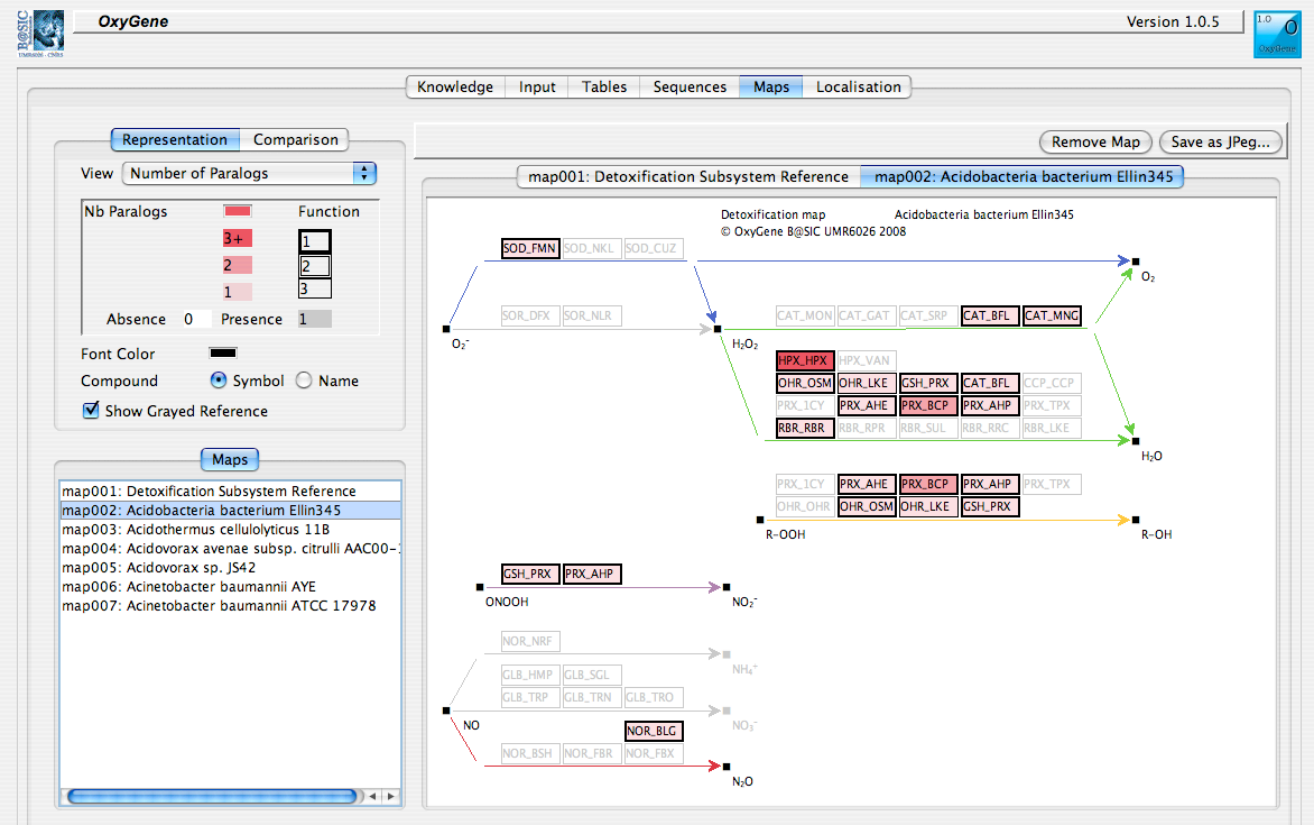
, then selecting the repertory and defining the file name. For example, the sequences shown above would be saved thus:

// choose different genomes same OxyDB

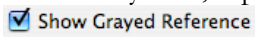
```
>BruAb1_0588|Brucella_abortus_9-941|SOD_FMN
MAFELPALPYDYDALAPFMSRETLEYHHDKHHQAYVTNGNKLLEGSGLEGKSFEEIVKESFGKNQALFNNAGQHYNHIHFWKWMKKDGGGKKLPGKLEKAFD
SDLGGYDKFRADFIAAGAGQFGSGWAWLSVKDGKLEISKTPNGENPLVHGAAPILGVDVWEHSYYIDYRNARPKYLEAFVDSLVNWDYVLEMYEKAA
>BruAb1_0933|Brucella_abortus_9-941|PRX_BCP
MAHPQVGDMAPDFTLPSDHGEITLSSLKGHPVVVYFYPKDDTSGCTREAIAFSQLKAEFDRIGVRVIGLSPDSATKHARFRTKHALTVDLVADEDRVALEAY
GVWVEKSMYGRKYMGVERTTFLIGADGRIAQVWNKVKVDGHAQAVLEAARRL
>BruAb2_0827|Brucella_abortus_9-941|CAT_MON
MTDRPIMTTSAGAPIPDNQNSLTAGERGPILMQDYQLIEKLSHQNRERIPERAVHAKGWGAYGTLTITGDISRYTKAKVLQPGAQTPMLARFSTVAGELGAA
DAERDVRGFALKFYTQEGNWDLVGNNTPVFFVRDPLKFPDFIHTQKRHPRTHLRSATAMWDFWSLSPESLHQVTILMSDRGLPTDVRHINGYGSHTYSFWND
AGERYWVKFHFKTMQGHKHWTNAEAEQVIGRTRESTQEDLFSAIENGEFPKWKVQVQIMPELDADKTPYNPFDLTKVWPHADYPPIDIGVMELNRNPENYFT
EVENAAFSPSNIVPGIGFSPDKMLQARIFSYADAHRHRLGTHYESIPVNQPKCPVHHYHRDGQMNVYGGIKTGNPDAYYEPNSFNGPVEQPSAKEPPLCISG
NADRYNHRIGNDDYSQPRALFNLFDAAQKQRLFSNIAAAMKGVPGFIVERQLGHFKLIHPEYEAGVRKALKDAHGYDANTIALNEKITAAE
>BruAb2_0930|Brucella_abortus_9-941|NOR_BSH
MKYQSQKVAMLYFYGALALFVAQVLFGVVAGTIYVLPNTLSVLLPFNIVRMIHTNALIVWLLMGFMGSTYYLLPEETETELYSTKLAVIQFWLFFVAAGVAV
AGYLFHIHEGREFLEQPFFIKVGIVVVCLIFLFNITLTALKGRKTTVTNILLFGLWGLALFFLFAFYNPINLALDKLYWWYVIHLWVEGVWELIMASILAFL
MIKLNGIDREVVEKWLYVIVGLALFSGILGTGHHYYWIGAPGYWQWIGSLFSTLEVAPFFTMVMFTFVMTWRAGREHPNRAALLWSIGCSVMAFFGAGVWGF
LHTLSSVNYYTHGTQLTAAHGHLAFFGAYVMLNLAAMAYAIPEIRGRTPYNQWLSMVSFWMMCTAMSVMTFALTFAGVVQVHLQRVLGENFMEVQDQLALFY
WIRLGSGVVVVISALMFVWAVLVPGRQRSQKLSGFAQQPAE
>BruAb2_0335|Brucella_abortus_9-941|GLB_TRC
MTILINQPHPSIDRDSIDRLVEIFYGRAREDEIIGPIFNRTVKDWDHHLARISEFWSSVILKTGGYDGRPMPPHLALNLENEHFDLWLELFEQTAQEIFPPE
AAIIFVDRARRIADSFEMAIATHSGRIRAPRHSRLPLIS
>BruAb2_0347|Brucella_abortus_9-941|OHR_OHR
MPILYTTQSTATGGRTGSAKTADGRLSVVLDTPKELGGQGGEGTNPEQLFASGYAACFLGALKFAAAKEKISIPAESTVTATVGIGPREDGTGFGLDVALSI
ALPGIDKAKAEELVQAAHIVCPYSHATRGNLDVRLSVA
>BruAb2_0527|Brucella_abortus_9-941|SOD_CUZ
MKSLFIASTMVLMAFPAFAESTTVKMYEALPTGPGKEVGTVVISEAPGGLHFKVNMEKLTPGYHGFHVHENPSCAPGEKDGKIVPALAAGGHYDPGNTHHHL
GPEGDGHMGDLPRLSANADGKVSETVVAPHLKKLAEIKQRSLMVHVGGDNYSDKPEPLGGGGARFACGVIE
>BruAb2_0522|Brucella_abortus_9-941|PRX_AHP
MLGIGDKLPSFKVTGVKPGFNHHEENGVSAFEEVTEQSFPGKWKVIFFYPKDFTFVCPTEIAEFARLASEFEDRDAVVLGGSTDNEFVKLAWRRDHKDLNKL
PIWSFADTNGSLVDGLGVRSPDGVAYRYTFVVDPDNVIQHVYATNLNVGRAPKDTLRVLDALQTDELCPCNREVGGETLKAA
```

3.5. The Maps tab

The *Maps* tab allows the biologist to view the subsystem (e.g. detoxification) for all submitted genomes. For example, the detoxification subsystem of a particular bacteria shows all detoxification genes the bacteria possesses within its genome, and how their corresponding enzymes are used to detoxify the oxidative compounds, such as O_2^- or H_2O_2. The Maps *tab* looks like this:



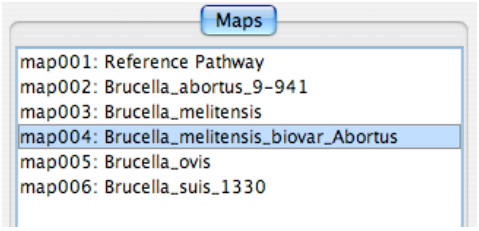
It includes a list of all available maps (the reference map + one map for every genome) appearing at the lower left corner of the client window, a *Representation/Comparison* panel just above (which will be described later), and finally a large frame on the right-hand side actually showing the map.

Note that the subsystem for a particular genome is a subset of the reference subsystem, which displays all enzymatic potentialities of all sequenced genomes within that oxidative stress domain. The reference subsystem is also stored as map001 and is already loaded as a sub-tab. It is the same reference subsystem as the one proposed in the *Knowledge* tab. The reference subsystem may also be shown on top of each specific bacteria subsystem, representing the absent genes and paths, by checking the *Show Grayed Reference* check-box  .

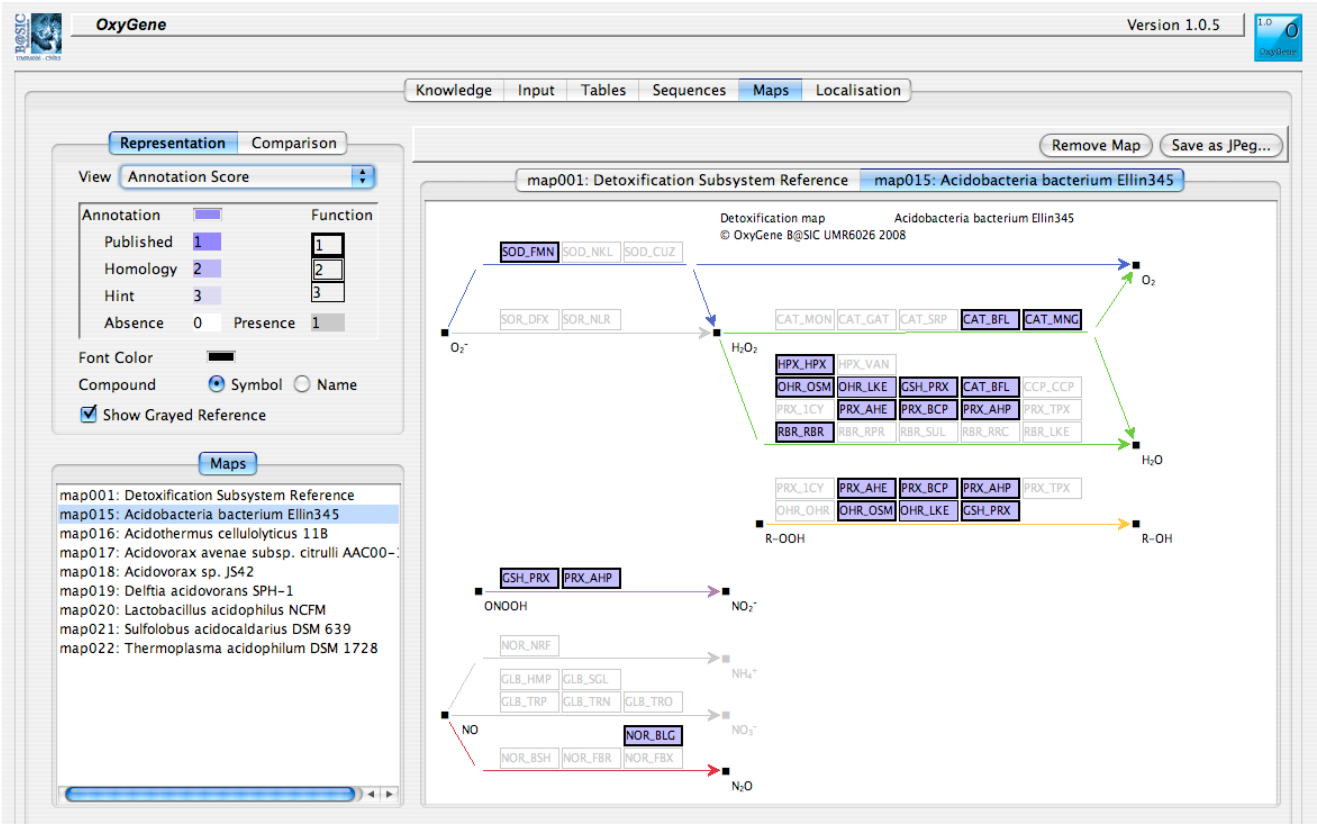
The OxyGene Client provides another interesting functionality, viz. the ability to construct new maps by combining old ones: the available operators are the intersection ∩ and union ∪ between any number of maps, and the difference δ between two maps. Thus it is possible for example to visualize what two bacteria have in common, e.g. the genes of detoxification they both possess, in which respect they differ, or whether they will be able to live together in some particular environment (e.g. do they each perform part of a particular detoxification process, thus enabling life where both bacteria alone could not have lived?).

3.5.1. Viewing Maps

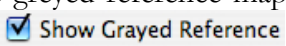
By clicking on a map name in the list of the maps appearing in the lower left corner of the client window,

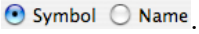


the biologist triggers the appearance of the corresponding map on the right hand side.

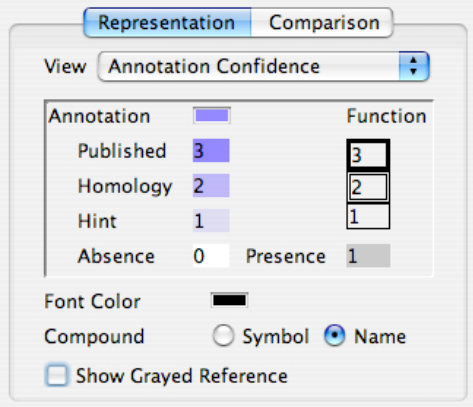


The map above represents the detoxification capabilities of the *Acidobacteria bacterium Ellin 345* bacteria, showing the enzymes whose corresponding genes are present within its genome, and displaying how they are used along the detoxification paths. In this case, for instance, the enzyme *SOD_FMN*, if expressed, may be used to detoxify the *Superoxide* compound into *Oxygen* and *Hydrogen Peroxide*, the latter being in turn detoxified by *CAT_MNG* into *Oxygen* and *Water*, or by *HPX_HPX* or *OHR_OSM* into *Water* alone.

The map above shows the greyed reference map underneath — this is obtained by checking the *Show Grayed Reference* check-box 

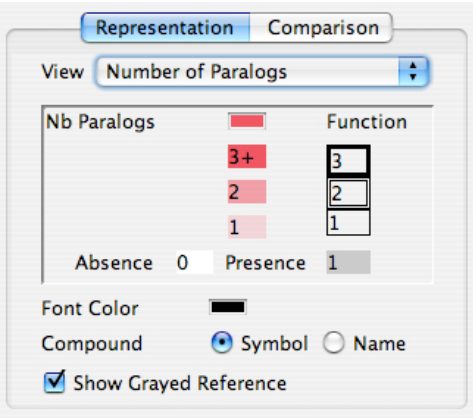
— and represents them using their chemical symbol rather than displaying their compound name — this effect is obtained by checking the appropriate *Symbol/Name* radio-button 

. The different colours used to draw the arrows represent different kinds of detoxification paths. The colour background of the enzymes corresponds to the annotation confidence level, as defined in the legend displayed on the *Representation* tab on the upper left side of the client window:



It is possible to change the font colour and the fundamental colour from which the different shades are derived by clicking on the corresponding small button . The frame thickness around the enzyme boxes represents the function confidence level, which is associated to each OxyDB gene *class*.

The biologist may want to view the number of paralogs rather than the annotation confidence level. This can be achieved by selecting *Number of Paralogs* within the *View* combo-box at the top of the panel. The map is then replaced by Figure 10, and the legend by:



The last colour, the *Presence* represented in grey, cannot be changed and is used in the domain reference maps and in combination maps (see below).

It is possible to remove a particular map by selecting it from the list in the lower left corner, and pressing the Remove Map button Remove Map. Likewise, to save a particular map under jpeg form, first select it from the same list, then adjust the colours, symbols and greyed items as desired, and eventually press the *Save as JPeg* button Save as JPeg...: a dialog window will appear, enabling the selection of the repertory and the definition of the file name.

At last, clicking on non-empty gene boxes triggers the opening of the corresponding gene window, as seen in Figure 10.

The maps in OxyGene were generated using the free Java library *JGraph.jar* that allows the production of graph visualization and layouts within Java programs (JGraph).

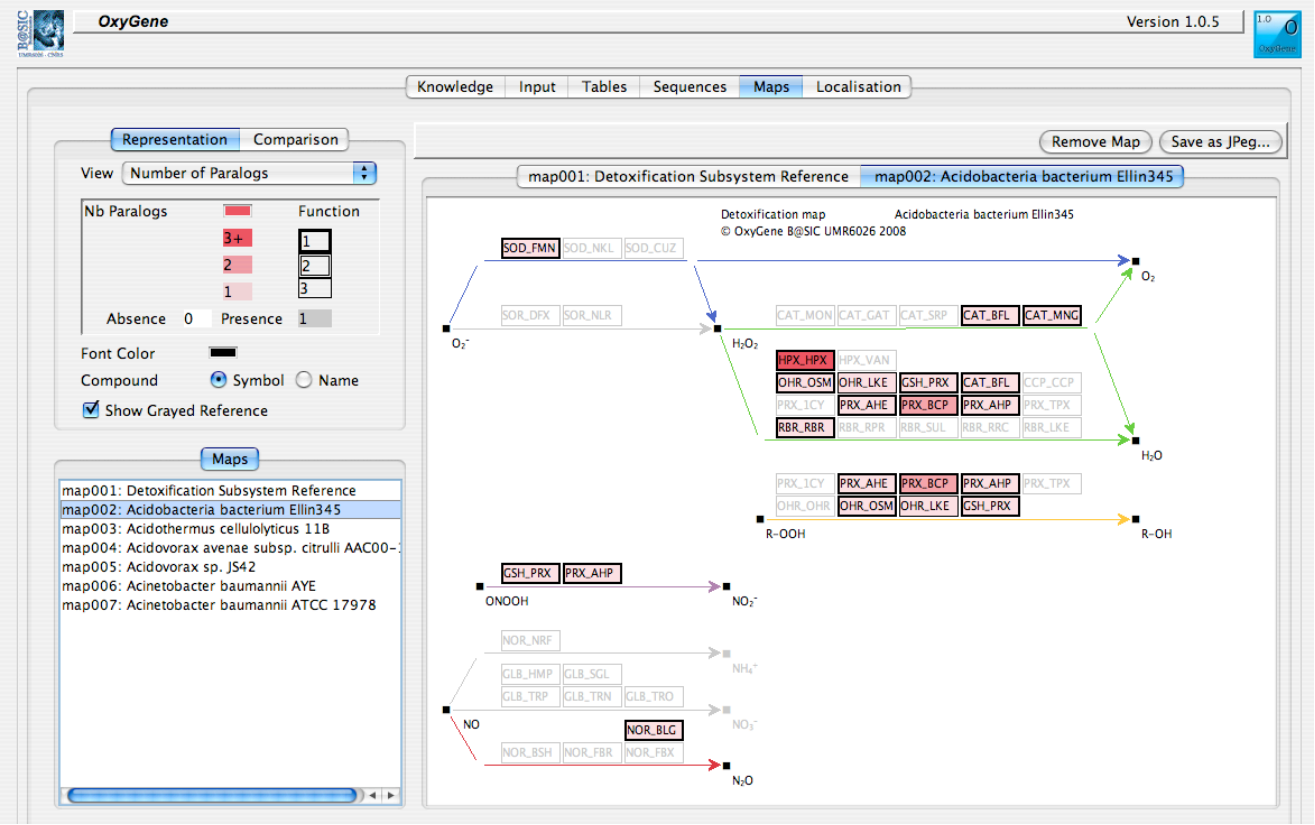


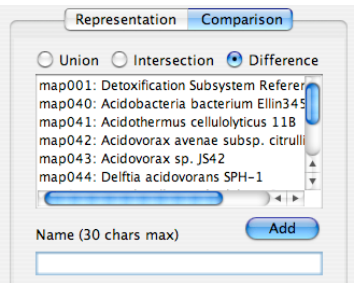
Figure 10. Map of *Acidobactereria bacterium Ellin345* representing the number of paralogs of each OxyDB gene class within its genome using different shades of the same red colour.

3.5.2. Comparing and combining Maps

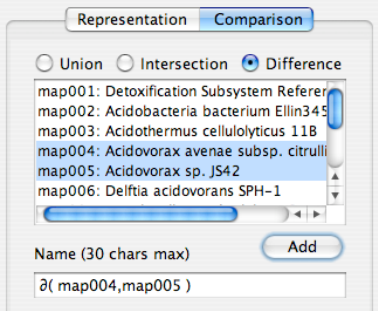
The OxyGene Client also allows the biologist to compare potentialities among organisms by combining their associated maps. Possible combinations are:

- **Intersection** of any number of maps: results in a map showing the presence of genes present in *all* original maps.
- **Union** of any number of maps: results in a map showing the presence of genes present in *at least one* of the original maps.
- **Difference** between exactly *two* maps: results in a map showing the presence of genes that are present in *one* of the two maps but *not in the other*.

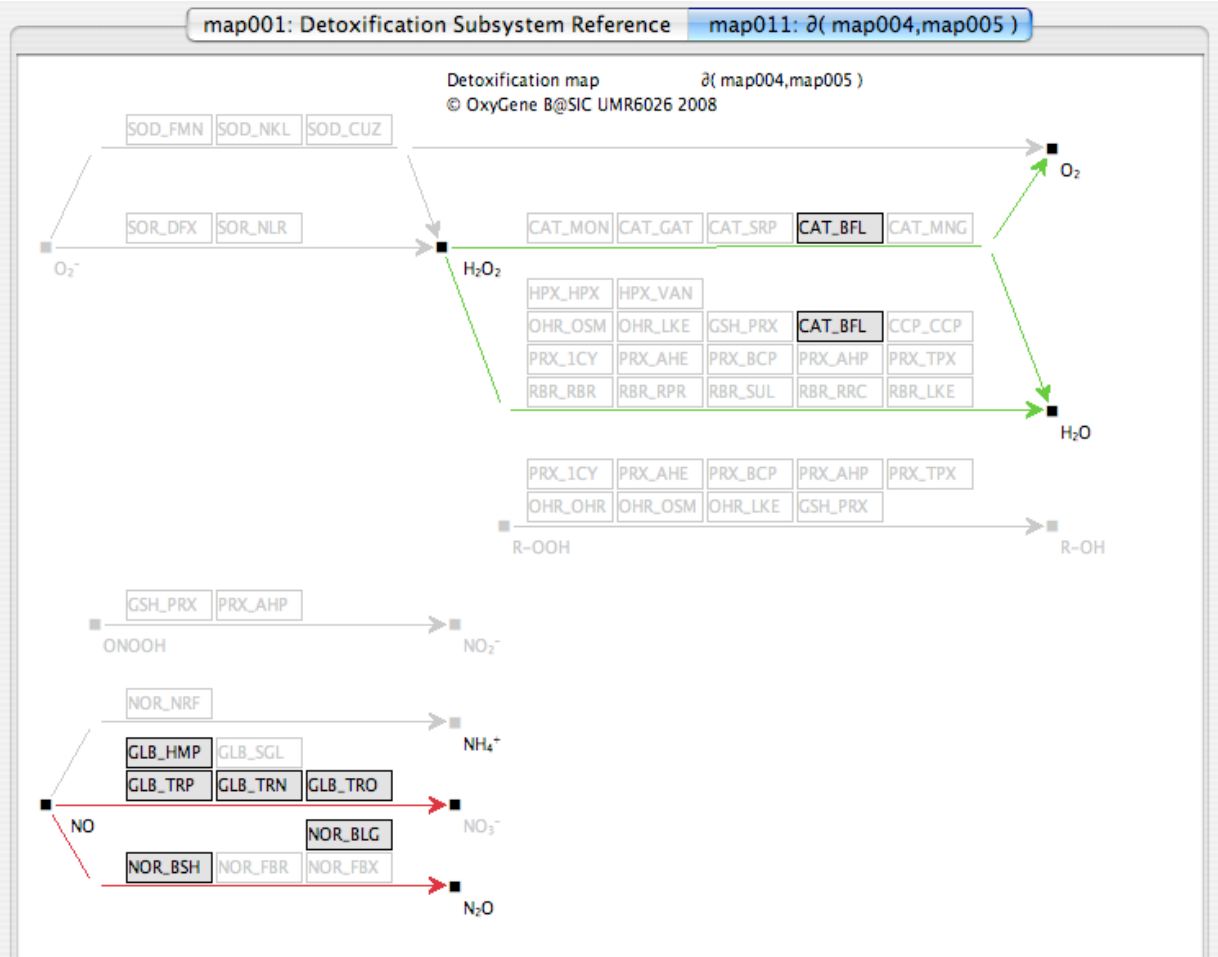
These operations are achieved by selecting the *Comparison* tab in the upper left panel of the client window. An appropriate interface appears:



It enables the user to easily select — in the usual manner — a number of different maps within the list provided (which contains all maps, including previous constructed combinations), to choose the desired operation by selecting the adequate radio-button, and finally to give a new name to the newly defined map. Then just press the *Add* button: the new map will appear on the right-hand side frame, and its name will be added to the lists of map names on the left-hand side of the client window. A "mathematical" name is supplied by default. For instance, constructing the difference between *Acidivorax avenae subsp. citrulli* and *Acidovorax sp. JS42* subsystems is achieved thus:



and produces the following map:

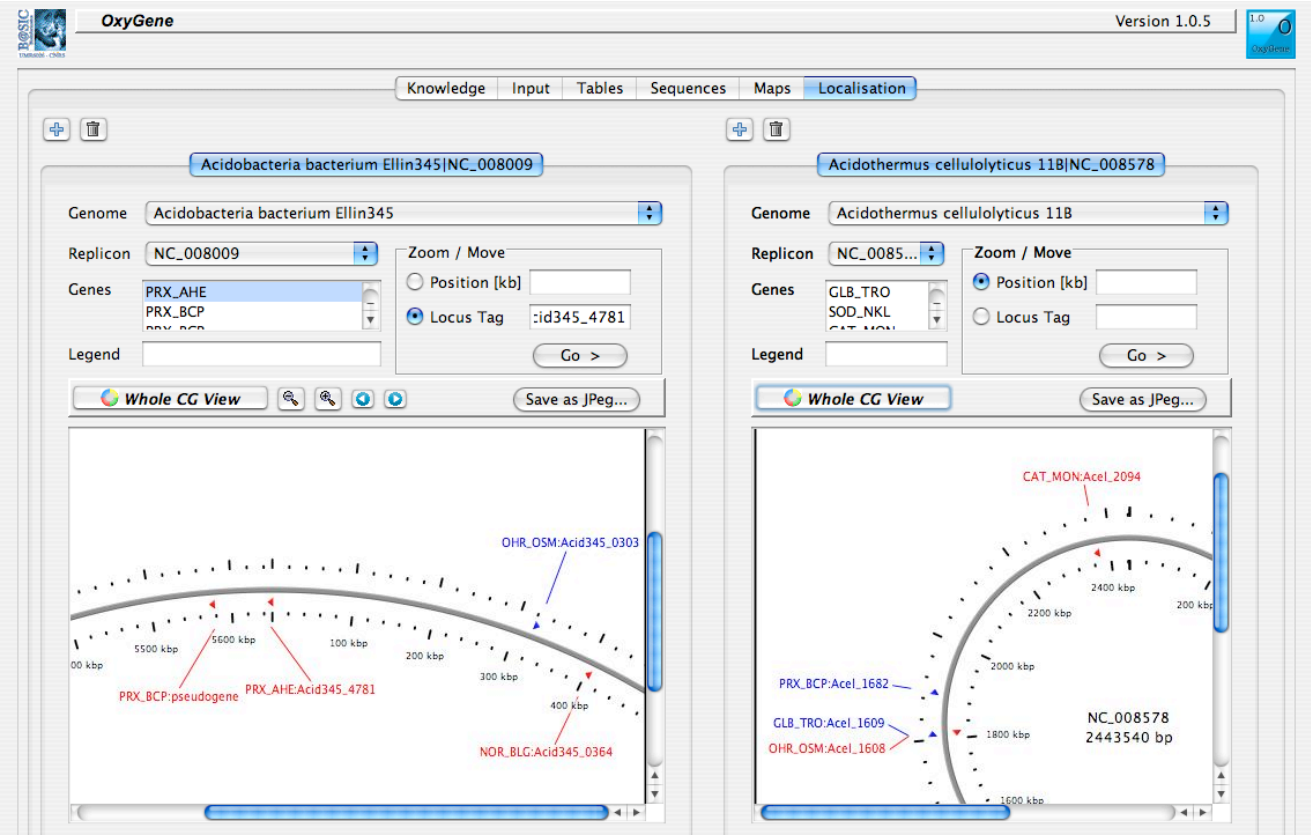


The presence of genes in the resulting combination maps is represented by a greyed background box around the gene name, whereas the absence — if displayed at all (recall *Show greyed Reference* check-box) — is denoted by a white — or just no — background as usual. Clicking on the gene boxes here does not trigger the appearance of the gene window, since it is unclear what genes should be associated to these logical operations.

3.6. The Localisation tab

The localisation tab is the last of the bio-analysis tabs provided. It enables the biologist to view the distribution of the OxyDB instances on the genome replicons. It makes use of the CGView Java library developed by Stothard P. and Wishart DS., supplied by the University of Alberta, Canada. The reference article is (Stothard & Wishart, 2005) and the web site is: `http://wishart.biology.ualberta.ca/cgview/`

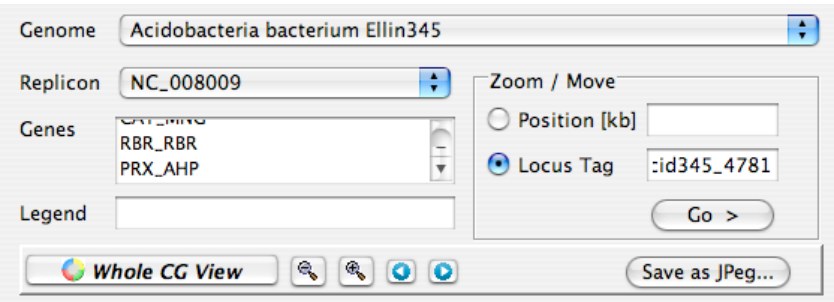
This tab is composed of two viewing frames in order to allow the comparison between different replicons. Let us take the example the two bacteria : *Acidivorax bacterium Ellin345* and *Acidothermus cellulolyticus 11B*.



The figures produced may help show how similar or dissimilar two bacteria or strains can be with regard to replicon localisation of detoxification genes.

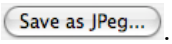
Adding a new tab to view a new replicon is achieved by clicking on the "+" button at the top of the client window on the left or right side, depending on where one wants to add the viewing tab. Likewise, clicking on the *dustbin* button removes the corresponding current viewing tab.

Inside the tab, the biologist must use the control panel:



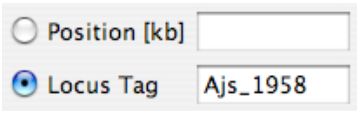
in order to display the desired replicon (or zoomed part of the replicon). The *Genome* combo-box at the top allows the user to select the desired genome; the *Replicon* combo-box underneath allows the selection of the replicon. The list of the OxyDB genes present within that replicon appears in the list box. The *Legend* field allows the biologist to add some commentary, which will be placed at the bottom of the figure. Clicking then on the *Whole CG View* button will generate the drawing of the whole selected replicon, with all the OxyDB genes associated.

The *CG View* then shows the whole circular replicon, with its name and size in the middle, its genome name at the top left corner, the positions in kbp (= kilo-base-pair) on the circular replicon, and finally the OxyDB genes drawn on their strand (forward in green and reverse in red) at their correct positions and with exact lengths. The labels used to identify the genes include the OxyDB name and locus tag.

The image drawn may be saved under jpeg form by clicking on the *Save as Jpeg* button 

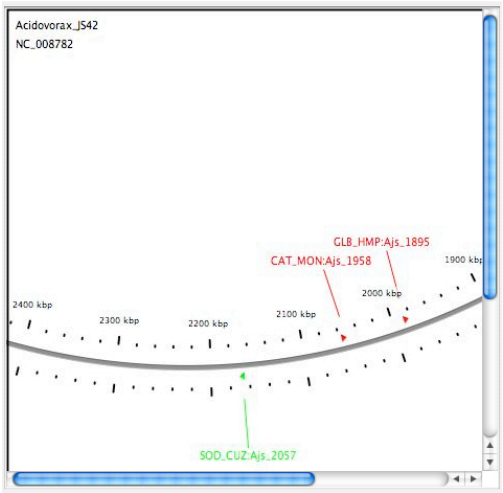
.

It is possible to zoom onto a particular position (expressed in kb) or onto a particular instance of OxyDB gene by using the *Zoom/Move* panel on the right-hand side of the control panel. To zoom onto a particular instance of an OxyDB gene, just click on the corresponding OxyDB name in the list box on the left. The *Locus Tag* radio-button will get selected, and the corresponding locus tag will be displayed in the associated editable field.



Just press the

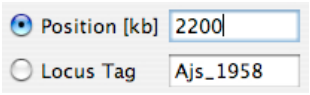
button then to display the CG View of the replicon centred onto that particular instance, e.g.



Then, clicking on the zoom buttons allows the bio-analyst to zoom respectively further inside or further away, the replicon window still being centred on the selected gene locus tag.

It is also possible to slightly rotate the replicon window; press the buttons to respectively rotate the window anticlockwise and clockwise.

Finally, use the *Position* radio-button and associated editable field to centre the replicon window onto a particular position, expressed in kbp.



Pressing then the *Go* button will result in displaying the replicon window centred on that position. The zoom and rotation buttons may further be used to view the replicon and OxyDB instances as desired.

4. Contact

We hope you will find OxyGene useful and this guide helpful.
If you have any questions or suggestions, feel free to contact us at: `fhubler@univ-rennes1.fr` or `stephane.avner@univ-rennes1.fr`.

Thank you,

The *B@SIC* team.

Bibliography

(Stothard & Wishart, 2005) Stothard P, Wishart DS. Circular genome visualization and exploration using CGView. *Bioinformatics* 21:537-539.

(Thybert & al., 2008) Thybert D., Avner S., Lucchetti-Miganeh C., Cheron A., Barloy-Hubler F. OxyGene: an innovative platform to investigate oxidative-response genes in prokaryotes whole genomes, submitted (2008).

(OxyDB, 2008) B@SIC, The OxyDB documentation (2008), `http://www.umr6026.univ-rennes1.fr/english/home/research/basic/software/`

(JGraph) `www.jgraph.com`.